Accurate, Efficient, and Insightful Quantum Chemistry Calculations of Non-Covalent Interactions: A Potential Method for Drug Design

Ka Un Lao

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210 USA

Abstract

One of the most promising ways to find drug candidates in drug design is to study the interactions between target biomolecules and drug candidates accurately. Although experimental screening methodologies keep improving, it is still costly and time-consuming to experimentally screen large numbers of potential compounds suitable to a target protein. On the other hand, computational screening technologies are an alternative method that can alleviate these challenges. However, these fast computational methods have lower accuracy than experimental approaches due to the approximations needed to make them computationally feasible. Hence, an efficient and accurate computational screening method for large numbers of potential compounds is in urgent need. For this reason, our group has recently developed a fragment-based quantum chemistry method called XSAPT (extended symmetry-adapted perturbation theory) to decompose the binding region (supersystem) between biomolecule and drug into biomolecular subsystems (fragments) in order to greatly reduce the computational cost without sacrificing accuracy. The main attractive feature of the XSAPT method is its ability to capture many-body polarization effects (important for systems with many fragments) which are omitted in traditional computational screening functions in drug design. Furthermore, the XSAPT interaction energy can be decomposed into physically meaningful energy components, and we can

explore how chemical modification of a potential drug molecule may change its binding affinity by studying the interplay of various energy components. From a computational point of view, XSAPT is "embarrassingly parallelizable", consisting of independent tasks that can be distributed across processors to reduce the scaling to only linear with respect to the number of fragments as opposed to the typical quantum mechanical methods which are at least cubic scaling with respect to the total system size. This fast theoretical method gives accurate binding energies for a variety of challenging non-covalent complexes, and these impressive results indicate that XSAPT is suitable for different binding environments. For example, XSAPT predicts a qualitatively-correct binding trend for a series of ionic-organic complexes as compared to experiment. Thus, XSAPT provides a route to understanding and controlling the ion-macromolecular binding property by modifying the structure of the macromolecule (a protein for example). The target of our XSAPT method is to predict accurate interactions between biomolecules and drug candidates. XSAPT has been employed in studying the interactions between an anti-cancer drug and DNA, and this binding complex consists of 157 atoms. A benchmark binding energy of -33.6 ± 0.9 kcal/mol is available from quantum Monte Carlo (QMC) calculations. Our XSAPT method yields a binding energy of -33.4 kcal/mol, within the statistical error bars of the QMC benchmark. Hence, accurate binding energies between DNA and drug molecules can be achieved by our XSAPT method. In summary, we demonstrated that XSAPT not only reduces the computational cost but also affords chemically-accurate interaction energies between molecules. These characteristics make XSAPT a promising method for use in fragment-based drug design to pre-screen large numbers of potential drug molecules.

Introduction

Pharmaceuticals are important to the quality of human life because they prevent and treat many diseases. In the drug design process, a compound is specifically designed

to interact with a specific biological target. The target might be an underexpressed or overexpressed molecule in the human body that causes disease, or is a biomolecule such as an enzyme originating from a disease-causing microorganism. For example, a specific drug molecule interacts with cholesterol to stop its absorption into the body or another drug might bind with an influenza protein to prevent the influenza virus from infecting new cells. In short, the search for compounds to effectively interact with these targets is the central step in drug design.

The most common way to develop a new drug is to screen a molecular library that includes large numbers of chemical compounds, and those compounds that can interact with targets receive further attention as potential drug candidates. This trial-and-error process can take many years and enormous amounts of money. Furthermore, the drug may not be effective enough or safe enough for human use at the end of the development period. Hence, a new and efficient method for drug development is demanded. Although some new experimental screening methods have been proposed recently, these methods still have many disadvantages, such as expensive instruments, the requirement of high sample concentration, and the difficulty of screening large numbers of compounds simultaneously.[1] Due to the rapid development of high-performance computers, computational screening technologies are alternative methods to design drug molecules. However, the binding energies predicted by these fast computational methods have lower accuracy than experimental approaches, because approximations have to be used to make them computationally feasible. Often, empirical scoring functions have only a loose connection to the real physics of intermolecular interactions are employed.[2,3] Hence, an efficient and also accurate computational screening method for large numbers of potential compounds is badly needed in the field of drug design.

It is essential to calculate interactions between molecules in the process of drug design by using accurate quantum mechanical (QM) methods which are more accurate than the traditionally empirical scoring functions. However, the computational cost of QM

methods that exhibit benchmark or "chemical" accuracy (errors $\lesssim$ 1 kcal/mol), such as CCSD(T), scales as $\mathcal{O}(N_s^7)$ where N_s measures the size of whole system. In other words, the computational cost increases 128 times when the system size doubles. The high computational cost limits its applicability to molecules with less than 30 atoms, which is not sufficient for computational drug design.

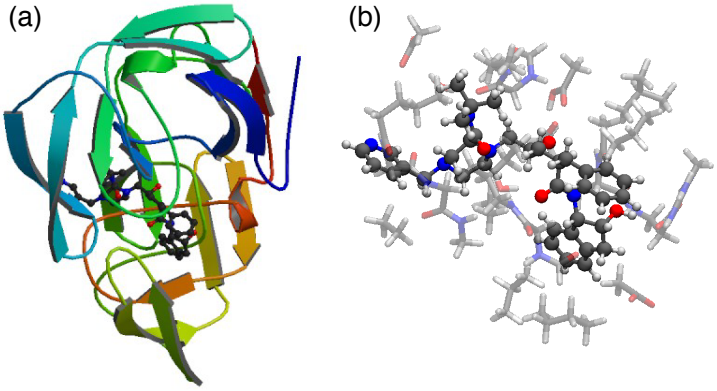

Figure 1: (a) Structure of the protease inhibitor indinavir bound to HIV protease, as obtained from PDB crystal structure 1HSG.[4] (b) An enlarged view of the binding pocket, consisting of indinavir (opaque ball-and-stick model) along with 16 amino acids and 2 crystallographic waters (translucent tubular model). [Panel (b) is reproduced from Ucisik *et al.*[5]; copyright 2011 American Institute of Physics.]

Consider the human immunodeficiency virus (HIV) protease + drug inhibitor system that is shown in Fig. 1(a). Even a representative model system (binding pocket of HIV protease + drug inhibitor), as shown in Fig. 1(b), contains 323 atoms including the drug indinavir (opaque ball-and-stick model) along with 16 amino acids and 2 crystallographic waters (translucent tubular model) which makes traditionally accurate QM methods are intractable. This work is aimed to build a rapid and accurate QM approach for calculating interactions between molecules to be applied to drug design.

Methods

Fragment-based quantum chemistry methods offer a way to surmount this predicament and rely on decomposing the large supersystem (the system including all fragments) into subsystems (fragments) to greatly reduce the computational cost.[6–8] Our group has

developed a fragment-based method called XSAPT[6,9–14] in an attempt to achieve chemical accuracy for interactions, yet remain affordable enough to be applied to systems such as the one in Fig. 1(b).

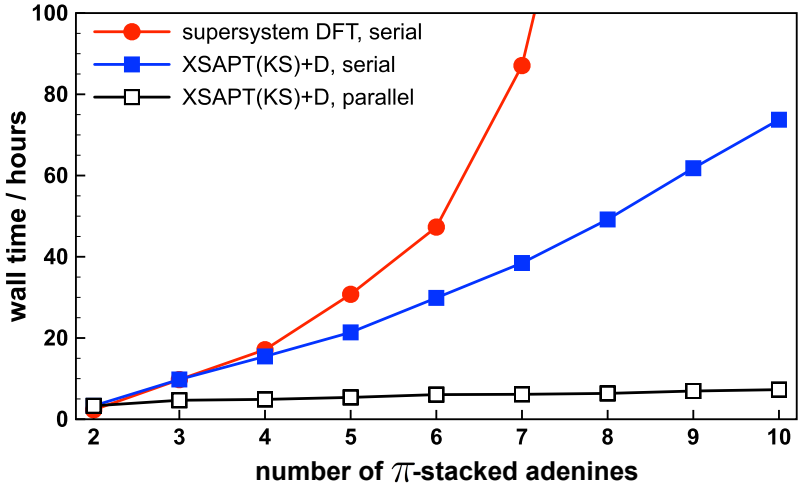

Figure 2: Timings for XSAPT(KS) and supersystem DFT calculations for π-stacked $(\text{adenine})_n$ systems. All calculations use the LRC-ωPBE functional and the hpTZVPP basis set.

I lead off with data illustrating the *efficiency* of the XSAPT method, as shown in Figure 2, which plots timings for XSAPT(KS)+D (a modified version of XSAPT). Plotted is the time required for a single-point energy calculation in $(\text{adenine})_n$ strands of increasing length (adenine is a DNA nucleobase). The supersystem DFT method is one of the most economical QM methods, and the timing results of this method are also shown in Figure 2 for comparison. Serial timings represent the total time required using one processor, and the scaling of XSAPT(KS)+D is $\mathcal{O}(n^2)$ where n is the number of adenine molecules. Parallel timings represent the actual elapsed time required when the calculation is simultaneously run on $n(n-1)/2$ processors for n adenine molecules. XSAPT is "embarrassingly parallelizable", consisting of independent tasks that can be distributed across processors to reduce the scaling to $\mathcal{O}(n)$. Even in serial, XSAPT(KS)+D is just as efficient as supersystem DFT for $(\text{adenine})_2$, and is substantially more efficient for larger systems. In parallel, the wall time required for an XSAPT(KS)+D calculation on

$(\text{adenine})_{10}$ is only about twice as large as that required for $(\text{adenine})_2$. Moreover, the XSAPT method is not only *efficient* as indicated in Figure 2, but also *accurate* as pointed out below.

Our XSAPT method is based on the combination of two fragmentation methods, an explicit polarization method (XPol) and symmetry-adapted perturbation theory (SAPT). This approach starts from the XPol method, and the primary function of this method is to capture many-body polarization effects (important for systems with many fragments) which are omitted in traditional computational screening functions in drug design. In a subsequent step, we apply a pairwise-additive form of SAPT to capture the rest of the interactions missing in the XPol step. However, the single-exchange approximation (SEA) used in the exchange formulas of SAPT makes SAPT unreliable for use in anionic systems.[13,15] Different levels of SAPT methods have been used to systematically study the binding energies in different ionic systems, and the rescaled formula for SEA has been proposed to resolve the problem in exchange formulas used SEA.[15,16] Such a rescaled formula for SEA can also been used in XSAPT. The resulting XSAPT method extends the traditional two-body SAPT method to many-body systems, and it maintains the computational cost of a two-body system. Furthermore, the XSAPT interaction energy can be decomposed into physically meaningful energy components,[12]

$$E_{\text{int}}^{\text{XSAPT}} = E_{\text{electrostatic}} + E_{\text{exchange}} + E_{\text{induction}} + E_{\text{dispersion}} \ , \tag{1}$$

and we can study the interplay of various energy components, some of which are attractive and some repulsive, to explore how chemical modification of a potential drug molecule may change its binding affinity. Furthermore, we have introduced many different techniques to improve the original XSAPT method:

[1] The use of Kohn-Sham (KS) orbitals for XSAPT, or XSAPT(KS). This incorporates intramolecular electron correlation beyond mean field theory in a relatively low-cost

way.[10]

[2] Long-range corrected (LRC) density functionals are used to obtain correct asymptotic behavior of exchange-correlation (XC) potentials in an automated and non-empirical way, instead of using traditional "splicing" schemes that simply "graft on" proper asymptotic behavior empirically. The latter approach is potentially problematic in the context of locating minimum-energy geometries.[13]

[3] Dispersion (van der Waals) interactions are problematic in XSAPT(KS), so we replace the dispersion terms in XSAPT(KS) with atom−atom dispersion potentials [XSAPT(KS)+D] that are fit to high-level QM data. Furthermore, I have reformulated XSAPT in the atomic orbital (AO) basis.[17] This formulation avoids the four-index integral transformation that is required in the original, molecular orbital version of the method.[9,11,14] This has the added benefit of reducing the scaling of XSAPT(KS) from $\mathcal{O}(N_f^5)$ to $\mathcal{O}(N_f^3)$ with respect to the fragment size, N_f.[10,12,14] Two XSAPT families of methods have also been proposed to achieve sub-kcal/mol binding accuracy. They are XSAPT+D with dispersion contraction [XSAPT+D+DC],[18] and XSAPT with coupled KS dispersion [XSAPT+CKS].[19]

Results and Discussion

Figure 3 shows binding energy errors for a variety of methods in a biologically-relevant dataset which includes a wide distribution of interactions in biochemistry. The XSAPT(KS) family of methods exhibit reasonably small errors and compete with other high-level methods that are vastly more expensive. All of the methods that outperform the XSAPT(KS) family of methods exhibit at least $\mathcal{O}(N_s^5)$ with respect to the size of the *whole system*, whereas XSAPT(KS) methods scale as $\mathcal{O}(N_f^3)$ with respect to size of a *fragment*.

The high specificity of binding of ions to a macromolecule (a protein for example) is highly important in the life cycle of cell, where ion binding often regulates enzymatic

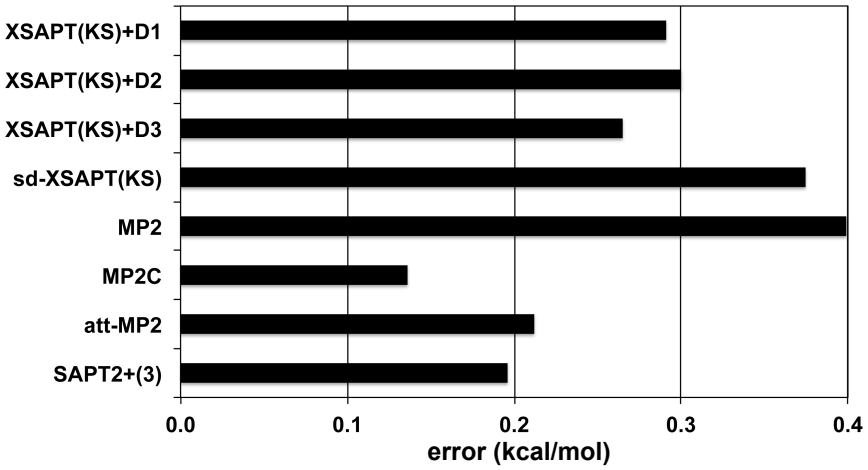

Figure 3: Mean absolute errors for (X)SAPT family of methods and the other supermolecular methods computed for S66 binding energies with respect to CCSD(T)/CBS benchmarks.

transformations. Hence, it is important to understand the ion-macromolecule binding property which provides a route to control the ion binding by modifying the structure of the macromolecule.

Figure 4 shows binding affinities versus experimental results[20] for a series of ionic organic complexes calculated by XSAPT(KS)+D and also the high-level CCSD(T) method, which is typically of benchmark quality but exhibits $\mathcal{O}(N_s^7)$ scaling. The errors in XSAPT(KS) and CCSD(T) are similar, and they also provide a qualitatively-correct binding trend for those ionic organic complexes as compared to experiment. Thus, XSAPT(KS) provides a key to understand the binding feature of ions and macromolecules.

The target of our XSAPT method is to predict accurate interactions between drug candidates and biomolecules. Here, we consider intercalation of the anti-cancer agent ellipticine into DNA, which involves insertion between two Watson-Crick CG base pairs, linked by their respective phosphate sugar puckers as depicted in Fig. 5. The structure depicted in the figure consists of 157 atoms, and a benchmark binding energy is available from quantum Monte Carlo (QMC) calculations which yield a binding energy of -33.6 ± 0.9 kcal/mol.[21] One of our family of XSAPT methods, sd-XSAPT(KS)[14] with

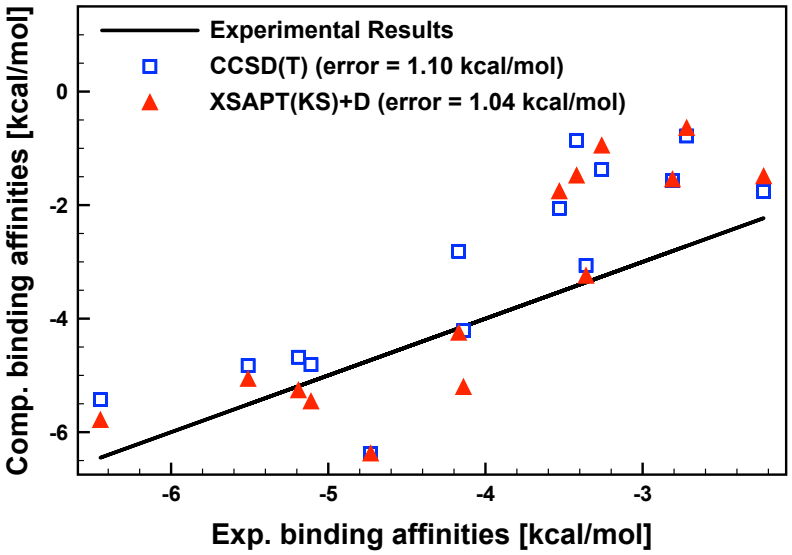

Figure 4: Binding affinities in kcal/mol versus experimental results[20] for a series of ionic organic complexes calculated by XSAPT(KS)+D and also the high-level CCSD(T) method.

a three-body dispersion correction, yields a binding energy of −34.4 kcal/mol, within the statistical error bars of the QMC benchmark. Thus, the accurate binding energy between DNA and drug molecule can be achieved. Furthermore, the XSAPT method has also been applied in molecular and ionic clusters, clathrates, and supramolecular complexes.[14,17,20]

Conclusion

In summary, we demonstrated that XSAPT not only reduces the computational cost (only linear scaling with respect to system size) but also affords chemically-accurate interaction energies between molecules. These characteristics make XSAPT a promising method for use in fragment-based drug design to pre-screen large numbers of potential drug molecules. In future work, we will use a low-cost classical molecular dynamics method to generate a series of reasonable binding positions between drug molecules and the binding pockets of biomolecules, and build a molecular binding library. Then, XSAPT can be used to predict the placement and binding energy of the drug biomolecule complexes in that library and compare the performance of traditional empirical scoring functions.

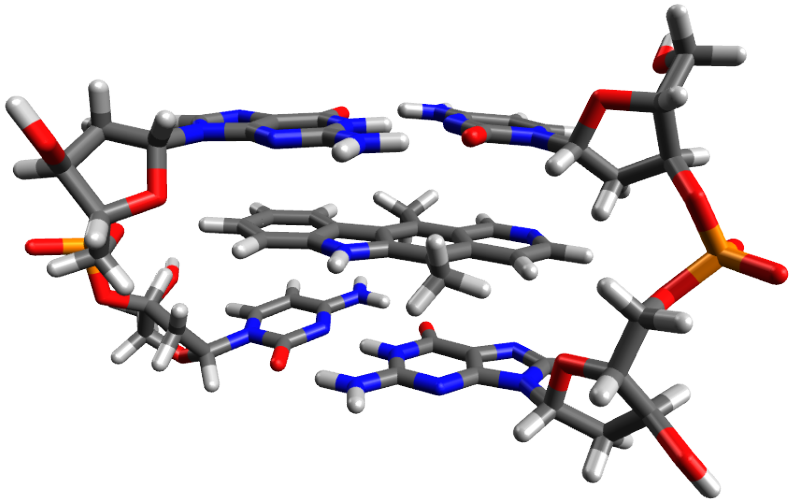

Figure 5: Ellipticine molecule intercalated into a GC:GC segment of DNA; binding energies computed with various methods are shown. XSAPT calculations used three fragments: neutral ellipticine and two single-stranded GC complexes, each with with a -1 charge.

Because XSAPT allows us not only to predict binding energies but also to understand their physical nature, systematic trends in the errors may emerge that provide clues as to how to improve empirical scoring functions.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550. Calculations were performed at the Ohio Supercomputer Center under project PAA-0003.

References

[1] Kumar, A.; Voet, A.; Zhang, K. Y. J. *Curr. Med. Chem.* **2012**, *19*, 5128–5147.

[2] Weill, N.; Therrien, E.; Campagna-Slater, V.; Moitessier, N. *Curr. Pharm. Design* **2014**, *20*, 3338–3359.

[3] Campagna-Slater, V.; Therrien, E.; Weill, N.; Moitessier, N. *Curr. Pharm. Design* **2014**, *20*, 3360–3372.

[4] Chen, Z.; Li, Y.; Chen, E.; Hall, D. L.; Darke, P. L.; Culberson, C.; Shafer, J. A.; Kuo, L. C. *J. Biol. Chem.* **1994**, *269*, 26344–26348.

[5] Ucisik, M. N.; Dashti, D. S.; Faver, J. C.; Merz, Jr., K. M. *J. Chem. Phys.* **2011**, *135*, 085101:1–12.

[6] Jacobson, L. D.; Richard, R. M.; Lao, K. U.; Herbert, J. M. *Annu. Rep. Comput. Chem.* **2013**, *9*, 25–56.

[7] Richard, R. M.; Lao, K. U.; Herbert, J. M. *Acc. Chem. Res.* **2014**, *47*, 2828–2836.

[8] Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632–672.

[9] Jacobson, L. D.; Herbert, J. M. *J. Chem. Phys.* **2011**, *134*, 094118:1–17.

[10] Lao, K. U.; Herbert, J. M. *J. Phys. Chem. Lett.* **2012**, *3*, 3241–3248.

[11] Herbert, J. M.; Jacobson, L. D.; Lao, K. U.; Rohrdanz, M. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7679–7699.

[12] Lao, K. U.; Herbert, J. M. *J. Chem. Phys.* **2013**, *139*, 034107:1–16.

[13] Lao, K. U.; Herbert, J. M. *J. Chem. Phys.* **2014**, *140*, 044108:1–8.

[14] Lao, K. U.; Herbert, J. M. *J. Phys. Chem. A* **2015**, *119*, 235–253.

[15] Lao, K. U.; Herbert, J. M. *J. Phys. Chem. A* **2012**, *116*, 3042–3047.

[16] Lao, K. U.; Schäffer, R.; Jansen, G.; Herbert, J. M. *J. Chem. Theory Comput.* **2015**, *11*, 2473–2486.

[17] Lao, K. U.; Herbert, J. M. **2016**.

[18] Lao, K. U.; Herbert, J. M. **2016**.

[19] Lao, K. U.; Herbert, J. M. **2016**.

[20] Lao, K. U.; Pomogaeva, A.; Chipman, D. M.; Herbert, J. M. **2016**.

[21] A, B.; Shulenburger, L.; Romero, N. A.; Kim, J.; von Lilienfeld, O. A. *J. Chem. Theory Comput.* **2014**, *10*, 3417–3422.